

Interactive comment on “Enabling BOINC in Infrastructure as a Service Cloud Systems” by Diego Montes et al.

A. Arribas (Referee)

alberto.arribas@informaticslab.co.uk

Received and published: 24 November 2016

General Comment: This paper describes the use of AWS infrastructure to run BOINC as an alternative to its traditional application using volunteer computing. The paper is significant as it addresses a timely issue for many research projects that, because of public interest and/or financial constraints, have chosen volunteer computing until now. The increased availability and reduced costs of many cloud platforms makes them extremely attractive for research projects and this paper should be attractive to large number of readers.

The approach, quality and presentation of the paper is good. Particularly, I would like to thank the authors for their effort on the Appendices to provide the reader with a clear step-by-step guide to reproduce their work.

C1

However, I believe the paper would benefit of further discussion of a series of points and I would like to encourage the authors to do so. Particularly, it would be extremely beneficial to have a better analysis of these points in the conclusions:

- Generating data is only one part of the story. Data needs to be stored and analysed. The costs of storage in the cloud can be substantial and the computational requirements for data analytics are very different to those needed to generate data in the first place, particularly if the data is to be open to a large number of scientists/data analysts

- Although I fully accept that this paper describes a proof of concept, the models run in this experiment are, relatively speaking, low resolution. There will be different challenges using higher resolution or more complex models. This could be related to the information shown in Figure 1 for run-time as it is surprising that smaller instances perform better than bigger instances. A better discussion of why this is the case would be very useful for other researchers as it has very clear financial implications.

Specific points: - Page 2, line 1: "the number of members in each ensemble tends to be small due to computational constraints" The use of computational resources in climate modelling is a balancing act between resolution, complexity and ensemble members as the authors point out a few lines earlier. It is not that the number of ensemble members is small due to computational constraints, it is often a conscious choice made by researchers. Having a big ensemble is desirable but it only makes sense if your model captures the right processes that you are interested in, otherwise you have the risk of just better sampling noise.

- Page 6, section 3.2.2 How could the data in S3 be used by other applications beyond climateprediction.net? How would this cope with much bigger datasets of hundreds of terabytes or petabytes?

A few typos have been detected. For example, in the opening pages: - page 1, line 15 should be: "... questions has led to them being ..." - page 1, line 17 should be: "... including climate modelling as a High ..." - page 2, line 24 should be: "... as well as

C2

proposed ..." - page 3, line 27 should be: "... For each instance's type at least 4 test ..."

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-193, 2016.